

Ruggedness Testing— Part II: Recognizing Interactions

Robert C. Paule, George Marinenko, Melissa Knoerdel, and William F. Koch

National Bureau of Standards, Gaithersburg, MD 20899

Accepted: August 28, 1985

This paper is a continuation of the preceding article which introduced the reader to the general concepts of ruggedness testing. The current paper describes the effects of interactions on the measurement process, and presents procedures for the separation of the main effects from the two-factor interactions. The general characteristics of interactions are described in some detail. A short-cut procedure is presented for the calculations. A number of examples of glass electrode measurements of pH of dilute acid solutions are used to illustrate ruggedness testing procedures.

Key words: glass electrodes; main effects; orthogonal designs; pH measurements; ruggedness testing; short-cut calculations; two-factor interactions.

Introduction

This paper is a continuation of the preceding (Part I) article which introduced the general principles of ruggedness testing. To be read in conjunction with Part I, it describes the effects of interactions on a measurement process and presents procedures for separating main effects and two-factor interactions.

Interactions and That Confounded Confounding

From Part I we know that an N measurement experiment can be used to determine $N-1$ main factors, provided the interactions are small. It is usually the case, in experiments involving well-behaved functions of the measurement variables, that when the main effects are small the associated interactions are very small. The

interactions are, in effect, the non-ideal departures from a simple additive model consisting of only constant main effects. Nevertheless, situations occasionally arise in which interactions are important.

In an eight-run, seven-factor experiment each main effect is confounded with 15 different possible interactions. Of the 15 interactions, the number and types are as follows: 3 two-factor, 4 three-factor, 4 four-factor, 3 five-factor, and 1 six-factor. Table 5,¹ which corresponds to the Yates-Youden design (see table 2 of part I), shows each of the main effects and the associated two- and three-factor interactions.

Table 5.¹ Interactions associated with the main effects.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>
<i>-BD</i>	<i>-AD</i>	<i>-AE</i>	<i>-AB</i>	<i>-AC</i>	<i>-AG</i>	<i>-AF</i>
<i>-CE</i>	<i>-CF</i>	<i>-BF</i>	<i>-CG</i>	<i>-BG</i>	<i>-BC</i>	<i>-BE</i>
<i>-FG</i>	<i>-EG</i>	<i>-DG</i>	<i>-EF</i>	<i>-DF</i>	<i>-DE</i>	<i>-CD</i>
<i>BCG</i>	<i>ACG</i>	<i>ABG</i>	<i>ACF</i>	<i>ABF</i>	<i>ABE</i>	<i>ABC</i>
<i>BEF</i>	<i>AEF</i>	<i>ADF</i>	<i>AEG</i>	<i>ADG</i>	<i>ACD</i>	<i>ADE</i>
<i>CDF</i>	<i>CDE</i>	<i>BDE</i>	<i>BCE</i>	<i>BCD</i>	<i>BDG</i>	<i>BDF</i>
<i>DEG</i>	<i>DFG</i>	<i>EFG</i>	<i>BFG</i>	<i>CFG</i>	<i>CEG</i>	<i>CEF</i>

About the Authors: Robert C. Paule is a physical scientist assigned to the NBS National Measurement Laboratory (NML). George Marinenko and William F. Koch are chemists in NML's Inorganic Analytical Research Division in which Melissa Knoerdel, a student, serves the Division during summer vacations.

¹ This paper (Part II), being a continuation of Part I, extends Part I's numbering system for tables and equations; its citations are listed in Part I's References section.

The Yates-Youden design (and the Plackett-Burman designs of a size such that $N=2^k$, where k is a positive integer) allow a relatively easy separation and determination of the more important confounding interactions. These designs allow one to use the Multiplication Rule for signs. The Multiplication Rule [4]¹ states that the pairwise multiplication of like signs produces a (+) and that of unlike signs produces a (-). Thus, looking at table 2 of Part I, the row pairwise multiplication of the signs for columns B and D produces the following column for the BD interaction:

BD
+
+
+
+
-
-
-
-

Note that this column is the exact opposite of the signs of column A , given in table 2. Thus, $-BD$ is the same as A . It is inseparable from A in the eight-run, seven-factor experiment since the values of the eight measurements are combined in an identical manner. Similar multiplications of signs shows that $A = -CE = -FG$. Multiplication of signs of the rows of columns " BC " and G produces the three-factor interaction BCG which is observed to be the same as factor A . Column " BC " can be simply obtained by using minus column F (see table 5). The confounding of all higher order interactions can be obtained by an extension of this general procedure.

If we wish to protect ourselves from misinterpretations due to large interactions, we must make more than N measurements for determining the $N-1$ main factors. To evaluate the main effects and all interactions, we must do the full factorial experiment. For seven factors this requires 128 measurements. Usually, however, one does not have to go this far. A reasonable compromise experiment consists of making two sets of N measurements which allow the separation of each of the main effects from the two-factor interactions. This compromise, however, does not separate among each of the two-factor interactions, and in addition it assumes that three-factor and other higher order interactions are unimportant. If we demand more information, then we have no choice but to make more measurements!

Let us again consider a seven factor pH experiment involving $2N (=16)$ measurements. This time we will use the previously reported first set of pH measurements

(see table 3 of Part I), and a third set of pH measurements which was made with all levels of the design reversed. Let us now consider the combined results from the first and third sets of measurements.

An examination of the signs of table 6, and the use of the Multiplication Rule, will show that the two-factor interactions $-BD$, $-CE$, and $-FG$ (which were grouped together in column 1 of table 5) still have an identical sign pattern but that this pattern is now different from the A main effect. The $-BD$ interaction has the following sign pattern $(- - - - + + + + - - - - + + + +)$. One can see that the last half of the interaction sign pattern is a repetition of the first half whenever an even-number of factors is multiplied together, but that the last half has a sign reversal whenever an odd-number of factors is multiplied together.

Table 6. Designs and test results.

Factor							Observed pH (milli-pH units)
A	B	C	D	E	F	G	
-	-	-	-	-	-	-	2904
-	-	+	-	+	+	+	3015
-	+	-	+	-	+	+	3006
-	+	+	+	+	-	-	2964
+	-	-	+	+	-	+	2999
+	-	+	+	-	+	-	3055
+	+	-	-	+	+	-	3049
+	+	+	-	-	-	+	2949
+	+	+	+	+	+	+	3040
+	+	-	+	-	-	-	2931
+	-	+	-	+	-	-	2978
+	-	-	-	-	+	+	3030
-	+	+	-	-	+	-	2967
-	+	-	-	+	-	+	2911
-	-	+	+	-	-	+	2874
-	-	-	+	+	+	-	2979

From table 6, and the Multiplication Rule, one can see that the three two-factor interactions within each column of table 5 are not separable from one another, but that they are separable from the main effects. The three-factor interactions are not separable from the main effects. A further consideration of table 6 will show that the main effects and their odd-numbered interactions are not separable from one another, but that they are separable from all of the even-numbered interactions. The nearest higher order interaction contamination for either the odd- or the even-numbered interactions is now two-factor multiples distant. If the magnitude of the interactions decreases as one goes toward higher order

interactions, then one has achieved a practical separation (isolation) of the main effects and of the groups of two-factor interactions.

Main effects A – G can be calculated from the data of table 6 by use of eq (1) (from Part I). The calculated respective effects are +51, –2, +4, +6, +27, +79, and –0.4 milli-pH units. The two-factor interactions are calculated in the same manner as the main effects. Note that the value of the “new N ” in eq (1) is the combined N from both sets of measurements (new $N=16$). As shown above, the sign pattern for the $-BD$ interaction is (– – – – + + + – – – – + + +). The value for the combined (BD , CE , FG) interactions is +11 milli-pH units. The other two-factor interactions can be calculated in a similar manner. Finally, we note from table 6 that if an offset had occurred between the first and third set of measurements, it would not affect the calculations of the main effects or the interactions. This immunity to offsets between the different sets of measurements is a consequence of using the Plackett-Burman based design. The PB-design will always have an equal number of positive and negative signs within each set so that the absolute level of the sets of measurements will not affect the calculations.

Short-Cut Calculations

All of the ruggedness testing calculations are conceptually quite simple, but are tedious to perform. Hand calculators that have at least nine memory registers allow short-cuts that minimize the arithmetic operations and the keying of the data. We will assume here that our calculations are made on sets of eight measurements. Let the average of these measurements be \bar{X} . Starting from eq (1) of Part I, the derivation of the short-cut method is as follows:

$$\begin{aligned} \text{Effect } A &= \frac{\sum A(+)}{N/2} - \frac{\sum A(-)}{N/2} \\ &= \frac{\sum A(+)}{N/2} + \left[\frac{\sum A(+)}{N/2} - \frac{\sum A(+)}{N/2} \right] - \frac{\sum A(-)}{N/2} \\ \text{Effect } A &= \frac{4\sum A(+)}{N} - 2\bar{X} \end{aligned} \quad (9)$$

For $N=8$,

$$\text{Effect } A = \sum A(+)/2 - 2\bar{X} \quad (9a)$$

Let us now rewrite the table 2 design of Part I for the first set of eight pH measurements, substituting the ordered measurement numbers for the positive signs.

For the set of eight measurements, one keys the measurements into memory registers 1–8, respectively, and then calculates the last term of eq (9a) which is two times the average of the eight measurements. This quantity is stored in memory register 9. In order to minimize the chance of error, it is advisable to use the measurement results that are stored in memory registers 1–8 to calculate this latter quantity. One then simply uses eq (9a) and the columns of table 7 to calculate the various effects:

$$\begin{aligned} \text{Effect } A &= (\text{Registers } 5+6+7+8)/2 - \text{Register } 9 \\ &= +40.75 \text{ milli-pH} \end{aligned}$$

$$\begin{aligned} \text{Effect } B &= (\text{Registers } 3+4+7+8)/2 - \text{Register } 9 \\ &= -1.25 \text{ milli-pH.} \end{aligned}$$

Table 7. Relabeled design and test results.

Factor							Observed pH
A	B	C	D	E	F	G	(milli-pH units)
—	—	—	—	—	—	—	2904
—	—	2	—	2	2	2	3015
—	3	—	3	—	3	3	3006
—	4	4	4	4	—	—	2964
5	—	—	5	5	—	5	2999
6	—	6	6	—	6	—	3055
7	7	—	—	7	7	—	3049
8	8	8	—	—	—	8	2949
Average							2992.625

The reverse sign PB-design listed in the bottom of table 6 can be similarly rewritten and used with eq (9a) to again calculate effects A – G .

Table 8 lists the calculated effects from the three sets of eight pH measurements that have been previously reported in tables 3 and 6. The actual, chronological order used for making our measurement sets consisted of the table 2 design of Part I, the reverse-sign design, the repeat table 2 design, and occasionally a repeat of the reverse-sign design. The labeling from our pre-

Table 8. Calculated effects.

Factor	(milli-pH units)						
	A	B	C	D	E	F	G
Set 1	41	–1	6	27	28	77	–1
Set 2	62	–3	2	–16	26	80	0
Set 3	48	–7	11	14	23	85	3

viously referenced data sets 2 and 3, will henceforth be reversed to conform to the chronological order. Thus, set 2 will now refer to the reverse-sign design, and set 3 to the repeat table 2 design.

We see that the set 2, reverse-sign PB-design, gives slightly different results. Note that the use of the Multiplication Rule on the reverse-sign PB-design (listed in

the bottom half of table 6), results in positive (rather than negative) two-factor interactions which are confounded with the main effects. For example, $A = +BD = +CE = +FG$. Let us take averages for the table 8 results for sets 1 and 2. For each of the averages of sets 1 and 2, the two-factor (and other even-number factor interactions) drop out. By similar reasoning the differences between sets 1 and 2, when divided by two, yields the separated even-number factor interactions. The set 1 and 2 averages and average differences are listed in table 9.

Table 9. Averages and average differences for effects.

Factor	For Data Sets 1 and 2 (milli-pH units)					
	A	B	C	D	E	F
Average	52	-2	4	6	27	79
Avg. Difference	11	-1	-2	-22	-1	2

The table 9 results for sets 1 and 2 are the same as the results obtained by the more tedious, direct calculations. The calculated effects from sets 2 and 3 could also be used to produce results comparable to table 9. It should be obvious that better (more stable) estimates can be obtained by first pre-averaging the effects from sets 1 and 3 before making the combined calculation with the set 2 data to produce the averages and average differences. The results from such calculations are given in table 10. For simplicity of presentation, these averages and average differences will hereafter be called the Main Effects and the Two-Factor Interactions, respectively.

Let us now summarize the short-cut calculations: for each data set use eq (9) or (9a), and its associated design table (such as table 7) to calculate the "contaminated" effects (as shown in table 8). Where possible, for like-sign designs, calculate the pre-averages. Also, for the like-sign designs calculate the squared differences of the "contaminated" effects (their use will be described in the next paragraph). The pre-averaged effects from the like- and reverse-sign designs are used to calculate aver-

ages and average differences (as shown in table 10). These latter averages and average differences are the separated, and relatively uncontaminated, main effects and interactions.

For PB-designs with $N=8$, the standard deviation of a single measurement is obtained by taking the square-root of the average of the above calculated squared differences of the "contaminated" effects (see eq (7) of Part I). For the current experiment s equals $\sqrt{384/7}$ or 7.4 milli-pH units. It has 7 degrees of freedom associated with it.

Judging the Main Effects and Interactions

To help decide if the main effects and two-factor interactions are real, or if they may simply be due to imprecisions in the measurements, let us once again use the t -statistic.

$$t = \frac{\text{effect under test}}{s_{\text{effect under test}}}$$

To determine the t -value we must evaluate the denominator of the equation. Since the main effects (ME) and the two-factor interactions (2FI) are calculated by taking either the averages or average differences from the same sets of data, the standard deviation of the ME and the 2FI will be the same. For the current example, the ME and the 2FI are calculated as follows:

$$\frac{1}{2} \left(\frac{\text{Sets 1+3}}{2} \pm \text{Set 2} \right)$$

The recognition of the form of the above calculations, and the use of the square of eq (2b), allow the evaluation of the standard deviation of the ME or the 2FI.

$$s(\text{ME or 2FI}) = \frac{1}{2} \left[\frac{4s^2/8 + 4s^2/8}{2^2} + \frac{4s^2/8}{1} \right]^{1/2} = \sqrt{3} s/4 \quad (10)$$

$$t = \frac{4 \times (\text{effect under test})}{\sqrt{3} s}$$

Table 10. Main effects and two-factor interaction.

Factor	For Data Sets [Avg. (1 and 3)] and 2 (milli-pH units)							Legend for + Factors
	A	B	C	D	E	F	G	
Main Effects	53	-4	5	2	26	81	1	A = 25 °C
Two-Factor Interactions	9	1	-3	-18	0	-1	-1	B = solution stirred
								C = 0.5 mL dilution
								D = 1 cm electrode immersion
								E = NaNO ₃ added
								F = KCl added
								G = pH measured at 10 min

For the current example s equals $\sqrt{384/7}$ or 7.4 milli-pH units. The observed t -value is as follows:

$$t_7 = 0.31 \times (\text{effect under test}).$$

For the main effect $A (=53)$ we have:

$$t_7 = 0.31 \times 53 = 16.$$

We see from table 10 that the main effects A , E , and F , and the two-factor interactions under columns A and D are statistically significant. This is very similar to our prior conclusions in Part I, except that we now distinguish between the main effects and the two-factor interactions.

Results From the Other Dilute Acid Solutions

A total of six different dilute acid solutions covering a pH range from 3.0 to 5.0 was tested. The purpose of these tests with the six solutions was to further evaluate the measurement procedures and to determine a practical upper limit for the pH measurements. Toward the end of the tests, the original glass electrode was broken and a second glass electrode of the same model and manufacturer was used as a replacement. This accidental breakage gave additional practical insight into ruggedness testing.

Some information regarding the six acid solutions is given in table 11.

The main design (table 2 of Part I) and the reverse-sign design were run with all six solutions. A standard pH=4.00 buffer was used to calibrate the pH meter before making each set of eight measurements. In many cases the designs were repeated and the standard deviations for a single measurement, s , were calculated. In accordance with the chronological order of the measurements, the main designs (table 2 of Part I) are labeled as sets 1 and 3, and the reverse-sign designs are labeled as sets 2 and 4. The calculated pre-averages for the like-sign designs are listed in table 12.

While making the measurements on sets 1, 2, and 3 of HCl solution 6, it was clear that there was great instability in the measurements. This instability is reflected in the

table-displayed standard deviation of 107 milli-pH units for solution 6. The listed pre-averages for solution 6 are also wild. The measuring equipment was operating properly. The problem appears to be associated with solution 6. Let us temporarily set the solution 6 results aside and examine the rest of table 12.

For glass electrode #1 there are rather large effects for the addition of either NaNO_3 or KCl (Factors E and F). The pre-averages appear to be larger for the H_2SO_4 solutions (1,2) than for the HCl solutions (3-5). For glass electrode #2, the pre-averages for NaNO_3 and/or KCl appear to be much smaller and are the same magnitude for both the H_2SO_4 and the HCl solutions.

The main effects and two-factor interactions were next calculated from the pre-averages by the short-cut procedures (see table 13). Average values for the standard deviation of a single measurement are also shown.

The general calculations described by eq (10), and the table-displayed s values, provide a method by which we can judge the main effects and the two-factor interactions. The standard deviations for either the ME or 2FI are 3.0, 3.2, 3.0, 6.5, and 7.4 milli-pH units for solutions 5, 3, 2, 1, and 4, respectively. From this we can conclude that an effect or interaction of more than 20 milli-pH units is not a chance happening.

Looking at table 13, we see that main effects A (lowering temperature), E (adding NaNO_3), and F (adding KCl) are large. The NaNO_3 main effect is about 30 for HCl solutions 5, 3, and 4, and is about 55 for H_2SO_4 solutions 2 and 1. The KCl ME is about 85 for HCl solutions 5 and 3, and 115 for the H_2SO_4 solutions. We note that the glass electrode #2 gives a KCl ME of only 18 for HCl solution 4. This result is not a fluke, but is the combined result from four separate experiments.

Two-factor interactions of appreciable size are observed under the column headings C , D , and G . Table 5 lists the possible 2FI interactions. It may be reasonable to assume that the 2FI associated with column C is due to the AE and/or BF interactions (temperature with NaNO_3 and/or stirring with KCl). It is known from experience that stirring can influence pH measurement

Table 11. Properties of acid solutions tested.

Label	pH	Acid	Ionic Contaminants	Electrodes used (#1 or #2) and 1984 Measurement Dates	
Soln. 5*	3.0	HCl	Hi purity water	#1 July 16-20	#2 Aug. 10
Soln. 3	3.7	"	" " "	#1 July 24-30	
Soln. 4	4.4	"	" " "	—	#2 July 31-Aug. 2
Soln. 6	5.0	"	" " "	#1 July 19-20	
Soln. 2	3.6	H_2SO_4	20 ppm by wt in water	#1 July 12-13	#2 Aug. 21
Soln. 1	4.3	H_2SO_4	3 ppm by wt in water	#1 July 9-12	#2 Aug. 20

* This solution was used for the above examples.

Table 12. Summary of pre-averages for main designs and reverse-sign designs.

Electrode	Soln.	Sets	pH	Milli-pH Units								Legend for + Factors
				<i>s</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	
Main Designs												
1	5	1+3	3.0	7	45	-4	9	21	26	81	1	<i>A</i> = 25 °C
1	3	1+3	3.7	8	35	-14	0	30	44	94	-15	<i>B</i> = solution stirred
1	6	1+3	5.0	107	-34	-44	-28	30	194	158	-98	<i>C</i> = 0.5 mL dilution
1	2	1+3	3.6	7	35	-15	1	44	60	120	-10	<i>D</i> = 1 cm electrode immersion
1	1	1+3	4.3	15	10	-21	-6	39	70	116	-18	<i>E</i> = NaNO ₃ added
												<i>F</i> = KCl added
2	4	1+3	4.4	14	36	-32	-4	-28	28	11	-32	<i>G</i> = pH measured at 10 minutes
Reverse-Sign Designs												
1	5	2	3.0	—	62	-3	2	-16	26	80	0	
1	3	2+4	3.7	10	40	-16	26	-22	28	86	6	
1	6	2	5.0	—	-35	-123	192	-32	70	45	134	
1	2	2	3.6	—	42	-9	23	-44	54	130	24	
1	1	2	4.3	—	17	-21	41	-31	45	90	28	
2	5	4	3.0	—	55	-6	11	-5	4	37	0	
2	4	2+4	4.4	27	29	-14	30	-2	19	26	8	
2	2	4	3.6	—	37	-7	12	-7	11	33	2	
2	1	4	4.3	—	27	-9	20	1	23	30	0	

Table 13. Summary of results—main effects and two-factor interactions.

Electrode	Soln.	pH	Acid	<i>s</i>	Milli-pH Units							Legend for + Factors
					<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	
Main Effects												
1	5	3.0	HCl	7	53	−4	5	2	26	81	1	<i>A</i> = 25 °C
1	3	3.7	"	9	38	−15	13	4	36	90	−4	<i>B</i> = solution stirred
1	2	3.6	H ₂ SO ₄	7	38	−12	12	0	57	125	7	<i>C</i> = 0.5 mL dilution
1	1	4.3	"	15	14	−21	18	4	58	103	5	<i>D</i> = 1 cm electrode immersion
												<i>E</i> = NaNO ₃ added
2	4	4.4	HCl	21	32	−23	13	−15	24	18	−12	<i>F</i> = KCl added
												<i>G</i> = pH measured at 10 minutes
Two-Factor Interactions												
1	5	3.0	HCl	7	9	1	−3	−18	0	−1	−1	
1	3	3.7	"	9	2	−1	13	−26	−8	−4	10	
1	2	3.6	H ₂ SO ₄	7	4	3	11	−44	−3	5	17	
1	1	4.3	"	15	4	0	24	−35	−12	−13	23	
2	4	4.4	HCl	21	−4	9	17	13	−4	8	20	

results. The temperature, NaNO_3 , and KCl are large effects, and it seems reasonable to assume that these effects may also be involved in the larger 2FI. By similar reasoning, the 2FI associated with column *G* may be due to *AF* and/or *BE* interactions (temperature with KCl and/or stirring with NaNO_3). The *D* column 2FI may be due to *AB* and/or *EF* interactions (temperature with stirring and/or NaNO_3 with KCl). It is interesting to note that the column *D* 2FI are appreciably different between the glass electrodes (#1 and #2), and also that the KCl ME with the HCl solutions is different between the two electrodes.

We were surprised by the slightly different characteristics exhibited by the two glass electrodes. In general, each laboratory will have its own set of surprises. The point to be made is that in ruggedness testing one should widely explore for factors that can influence the measurement results. Conversely, it seems wise not to be too concerned with determining higher order interactions, or with making extremely precise determinations of the main effects and two-factor interactions.

Can a factor such as the two different glass electrodes be used as a factor in the PB-design? It is certainly not a quantitative factor, such as temperature, and we cannot set the electrodes at a high or a low level. Nevertheless, we can accept the two levels we happen to get, and can go through all of the calculations. The advantages of orthogonal effects are maintained. With 100% hindsight, we see that it would have been wise to include the two different glass electrodes as one of the factors in the PB-design. As it turned out, we stumbled onto this factor by making a sequential series of PB-experiments.

How should we react to the observation of an important nonquantitative factor such as the glass elec-

trode? We cannot easily tighten the glass electrode specifications. Clearly, further exploratory work dealing with the basic science of the glass electrode measurement process is needed.

As a final note we present table 14 which suggests an upper limit for precise pH measurements of a fully dissociated acid such as HCl . The results shown in table 14, which were collected from previously listed tables, suggest an upper pH limit of about 4.5. Higher pH measurements, associated with solutions 6, are plagued by the solution's absorption of CO_2 from the atmosphere. Special measurement procedures are required for pH measurements above 4.5.

Table 14. Effect of pH on standard deviation.

pH	<i>s</i> (milli-pH units)
3.0	7
3.6	7
3.7	9
4.3	15
4.4	22
5.0	107

Conclusions

The principles of ruggedness testing involving the recognition of interactions have been discussed and illustrated by a study of pH measurements using glass electrodes. Separation of the main effects and the two-factor interactions was obtained. The statistical significance of individual main effects, and of groups of two-factor interactions was determined. The precision of the pH measurements was found to be a strong function of the pH level. The physical interpretation of these findings is further discussed in companion articles, [5] in particular, of the *NBS Journal of Research*.